Supplementary Information

Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges

Vinayak Agarwal,[1,8] Jessica M. Blanton,[1,2,8] Sheila Podell,[1,2] Arnaud Taton,[3] Michelle A. Schorn,[1,2] Julia Busch,[1,2] Zhenjian Lin,[4] Eric W. Schmidt,[4] Paul R. Jensen,[1,2] Valerie J. Paul,[5] Jason S. Biggs,[6] James W. Golden,[3] Eric E. Allen,[1,2,3*] and Bradley S. Moore[1,2,7*]

[1]Center for Oceans and Human Health, Scripps Institution of Oceanography, University of California, San Diego

[2]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego

[3]Division of Biological Sciences, University of California, San Diego

[4]Department of Medicinal Chemistry, University of Utah, Salt Lake City

[5]Smithsonian Marine Station at Ft Pierce, Florida

[6]University of Guam Marine Laboratory, UoG Station, Guam

[7]Skaggs School for Pharmacy and Pharmaceutical Sciences, University of California, San Diego

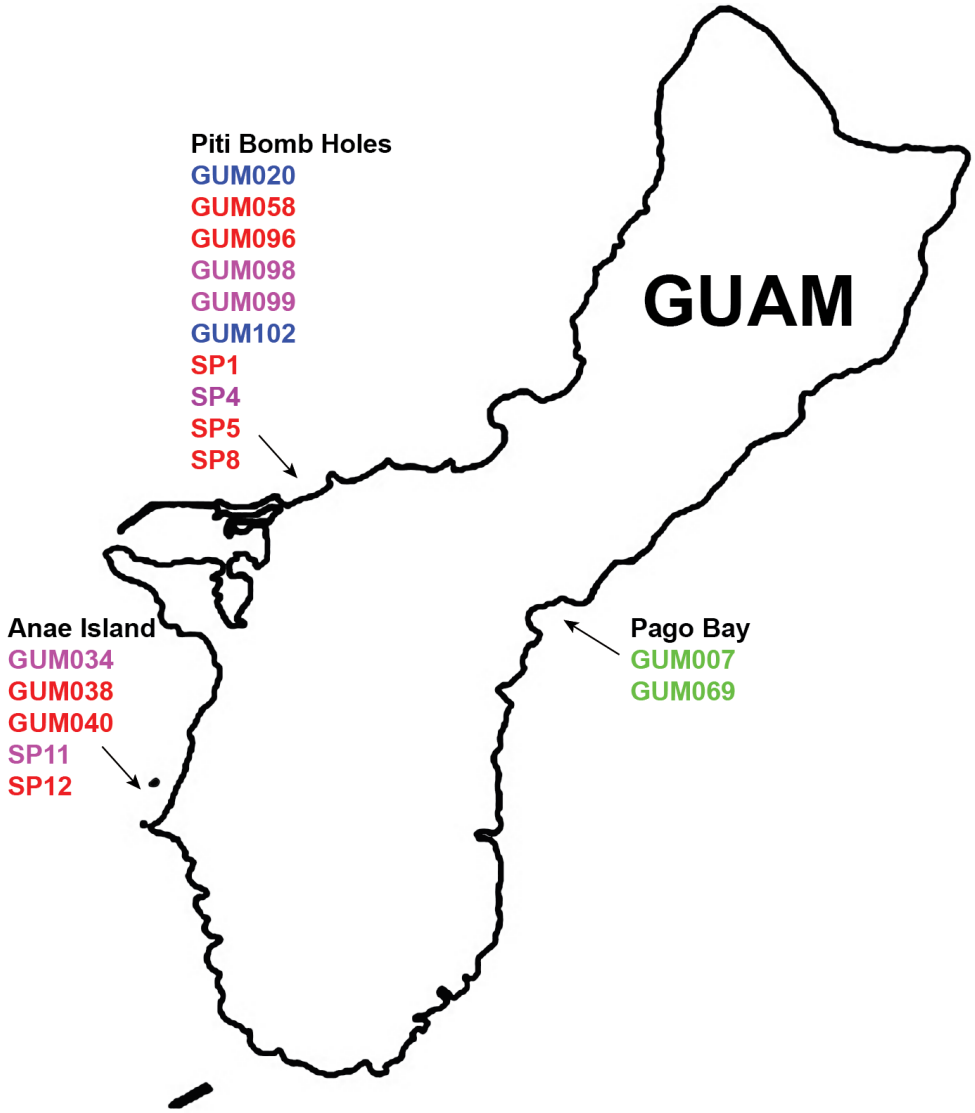
[8]These authors contributed equally to this work

*To whom correspondence should be addressed: bsmoore@ucsd.edu or eallen@ucsd.edu

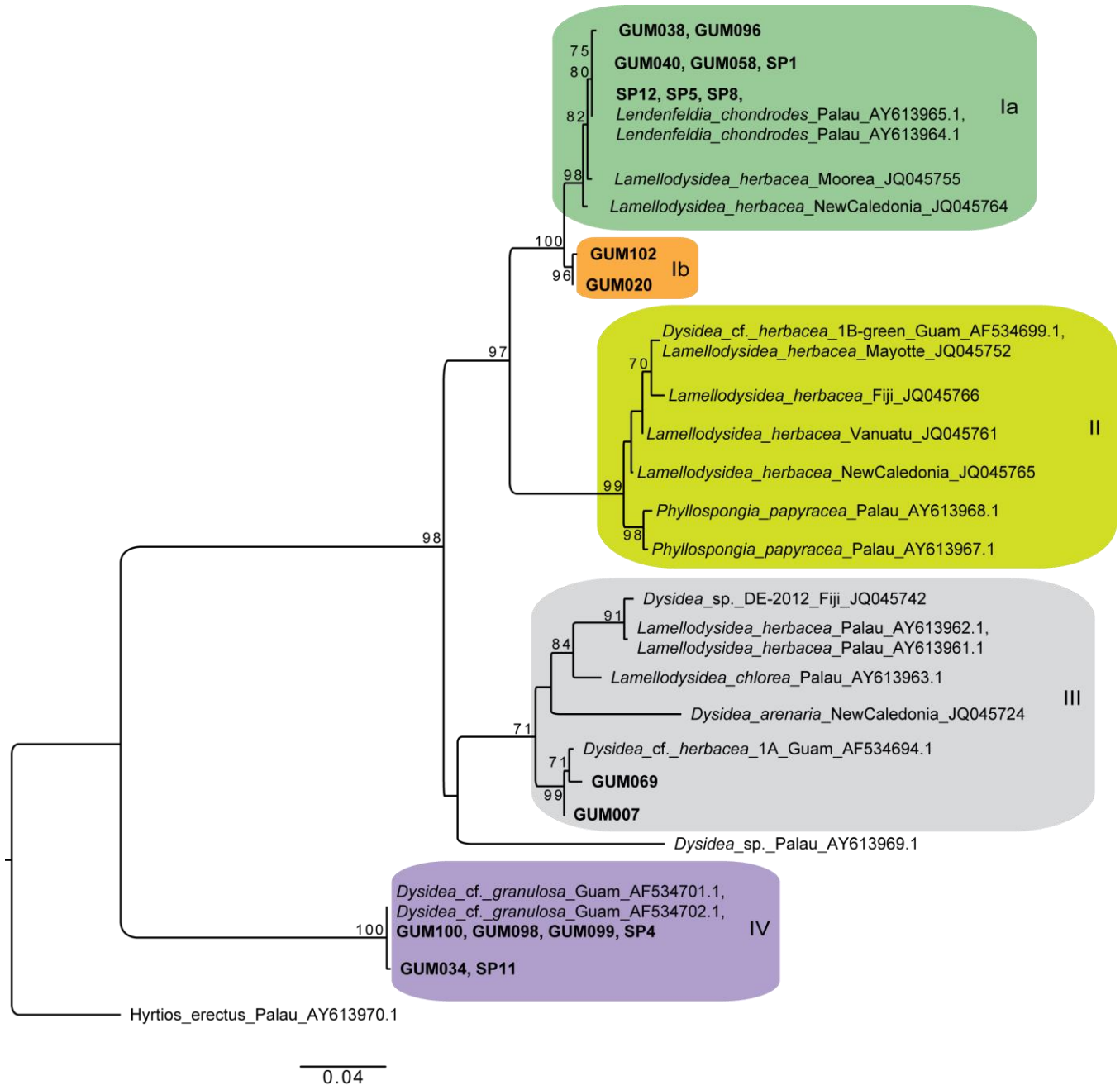
Table of contents

Supplementary Results

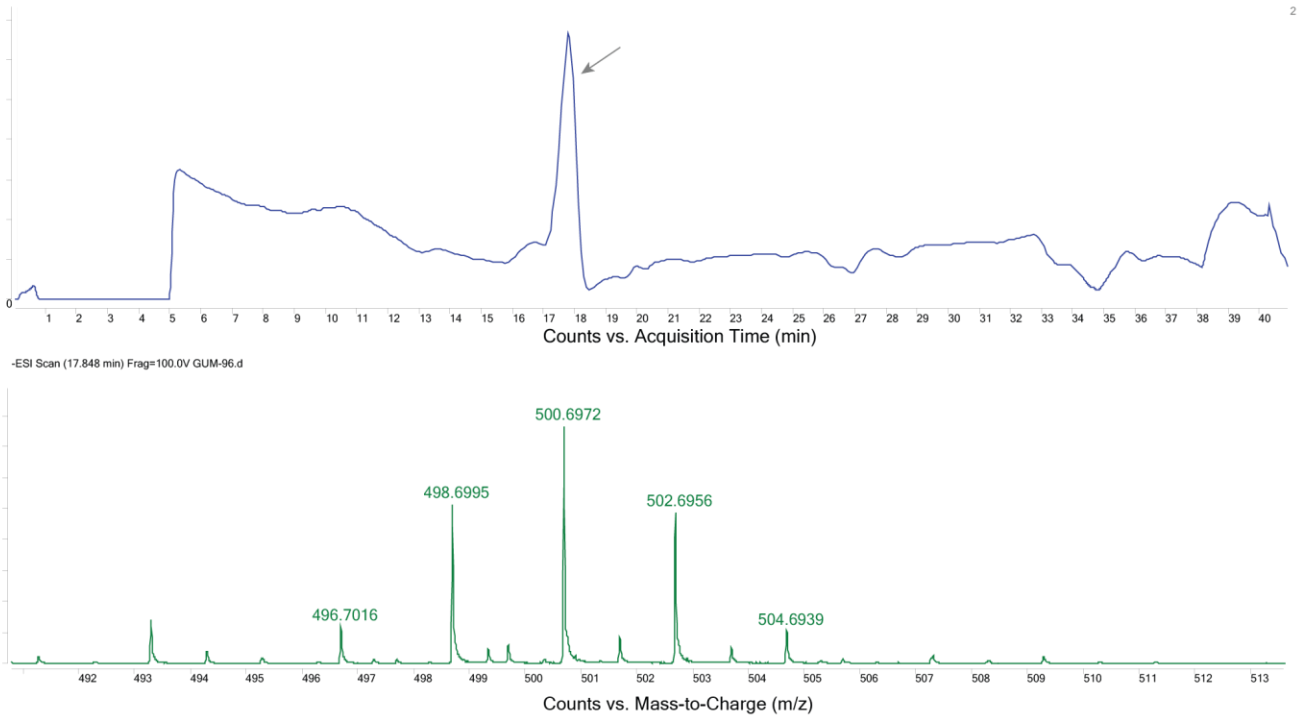
Supplementary Figures



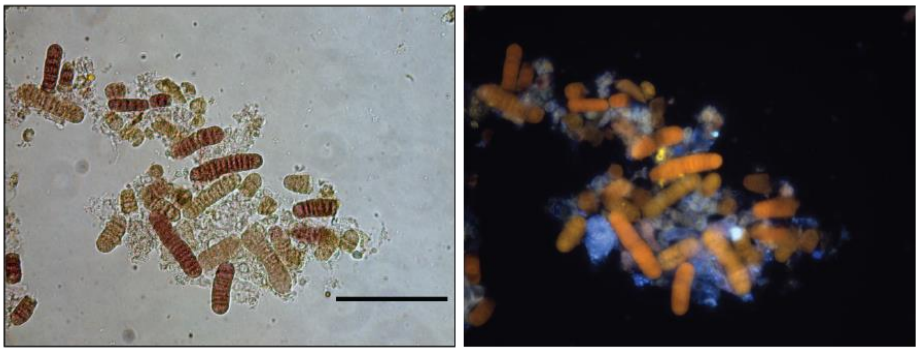
Supplementary Figure 1: Collection sites in Guam for sponge specimens used in this study. Sponge specimens are color coded according to their phylogenetic placement in Figure 2A with sponges placed in Clade Ia colored red, sponges placed in Clade Ib colored blue, sponges placed in Clade III colored in green, and sponges placed in Clade IV colored in purple. No sponges belonging to Clade II could be collected from Guam as a part of this study. All sponge specimens SP* were collected in 2014, specimens GUM* were collected in 2015.



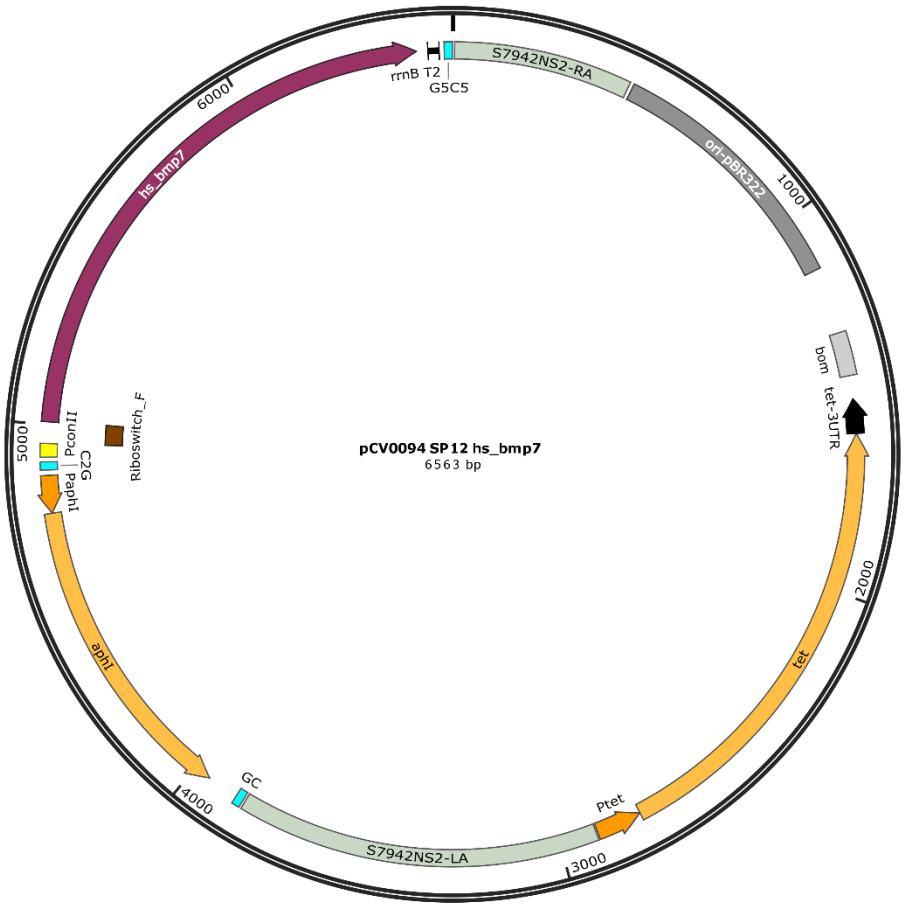
Supplementary Figure 2: Expanded sponge maximum likelihood phylogeny based on ITS-2 gene sequences, rooted on *Hyrtios erectus*. Included sequences are from specimens from the current study as well as additional sequences from the literature ranging multiple Indo-Pacific locations and reflecting inconsistent taxonomic identifications. Bootstrap values ≥ 70 are shown, scale bar indicates substitutions per site, and identical sequences are listed at a single branch tip.



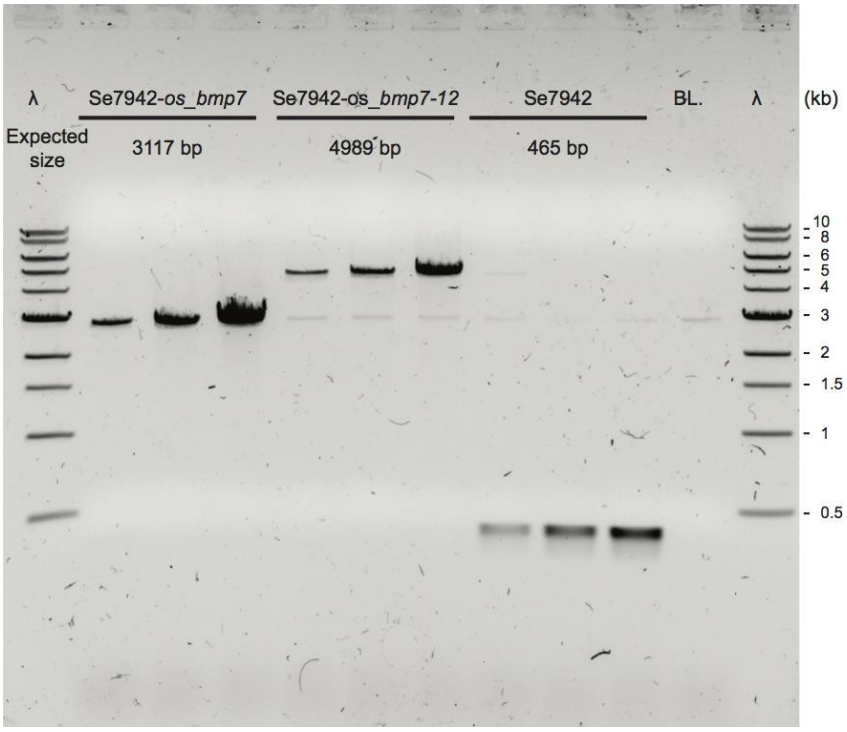
Supplementary Figure 3: Detection of **1** by LC/MS/MS in Clade Ia sponge specimen GUM096 organic extract. The top panel shows an extracted ion chromatogram at MS1 m/z = 500.69; 10 ppm accuracy. The $[M-H]^{1-}$ MS1 spectra corresponding to the marked peak is shown in the bottom panel which corresponds to the chemical formula of **1**, consistent with prior literature reports describing the co-isolation of **1** and **5** from Dysideidae sponges[1-4].



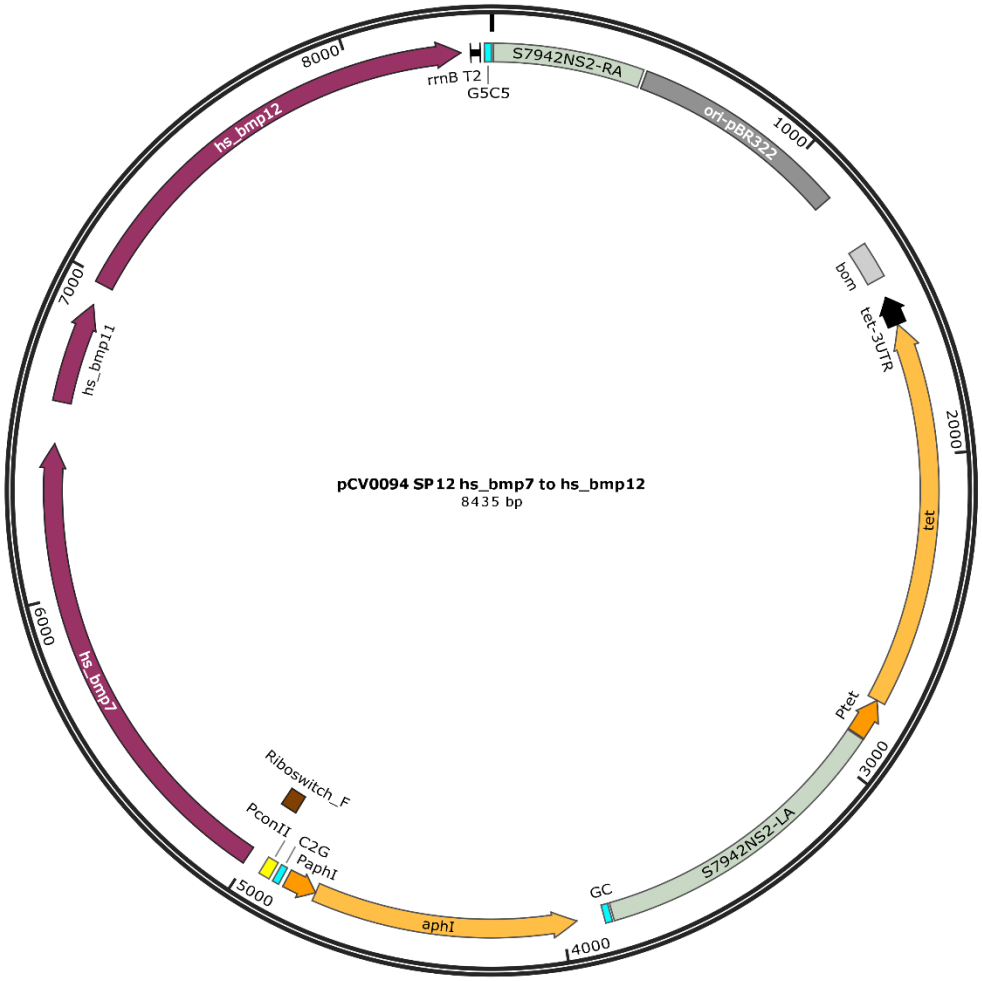
Supplementary Figure 4: Morphology of cyanobacterial cells from Clade III GUM007 tissue homogenate. Cyanobacterial cells and characteristic photosynthetic pigment autofluorescence of *H. spongeliae* are seen as red and yellow filaments under light (left), and fluorescence (right) microscopy. Scale bar represents 50 µm.



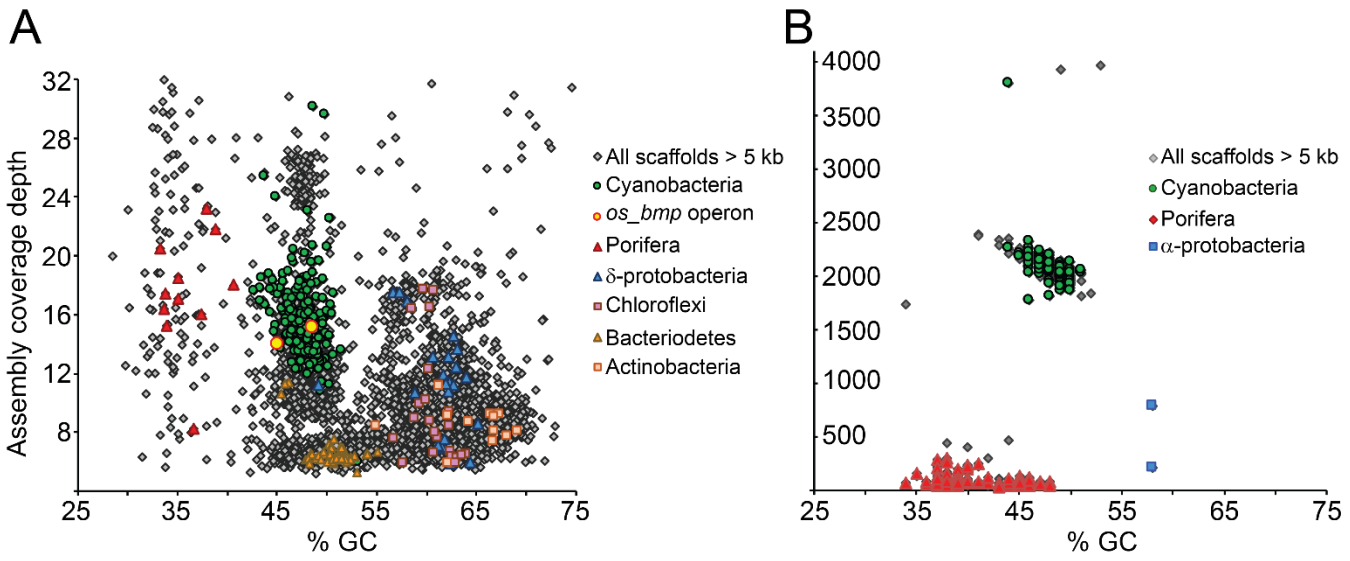
Supplementary Figure 5: Vector map of plasmid pCV0094_SP12_*hs_bmp7*. The backbone of this plasmid, pCV0094 was assembled from 3 modular devices obtained through restriction digests using either ZraI or EcoRV from the donor plasmids pCVD023, pCVD003 and pAM5019. pCVD023 carries the *S. elongatus* NS2 homologous regions, a tetracycline resistance marker, the *E. coli* pBR322 origin of replication and bom site; pCVD003 carries the kanamycin resistance marker *aph1*; and pAM5019 carries a gene expression cassette harboring the promoter P_{conII}, the riboswitch F, a cloning site and a terminator sequence. The combinatorial construction of pCV0094 was performed using the GeneArt Seamless Cloning kit and relied on short 21-nucleotide G- and C-rich overlapping sequences to guide the combination of the different modules in a specific order and orientation.



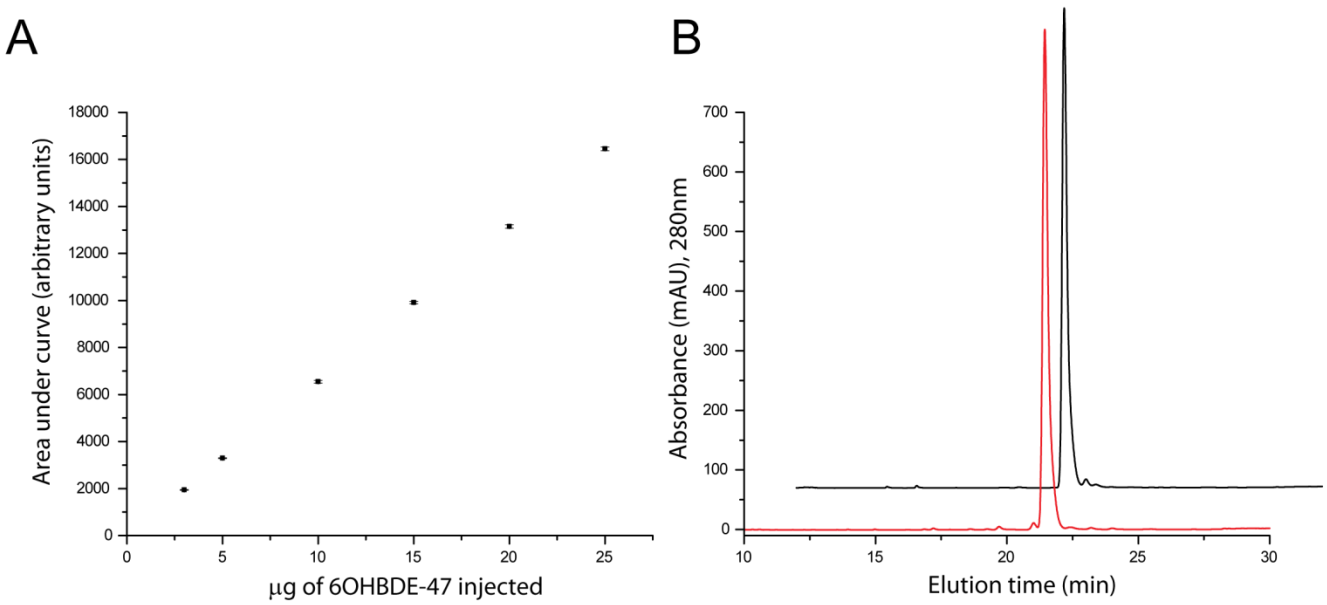
Supplementary Figure 6: PCR analyses carried out on *S. elongatus* strains (Supplementary Table 6) to confirm chromosomal integration of the recombinant DNA and complete segregation of the engineered chromosomes. PCR products of the expected size were obtained for the recombinant (*Se*7942-*hs_bmp7* and *Se*7942-hs_bm*p7-12*) and WT (*Se*7942) strains, while there was not any WT band for the recombinant strains. For each strain, three PCR reactions with increasing amount of template cultures were performed. A 6 kb light background band was seen in the first lane of Se7942 and a light background band at about 3 kb was seen in all lanes including the blank (BL) PCR reaction without template.



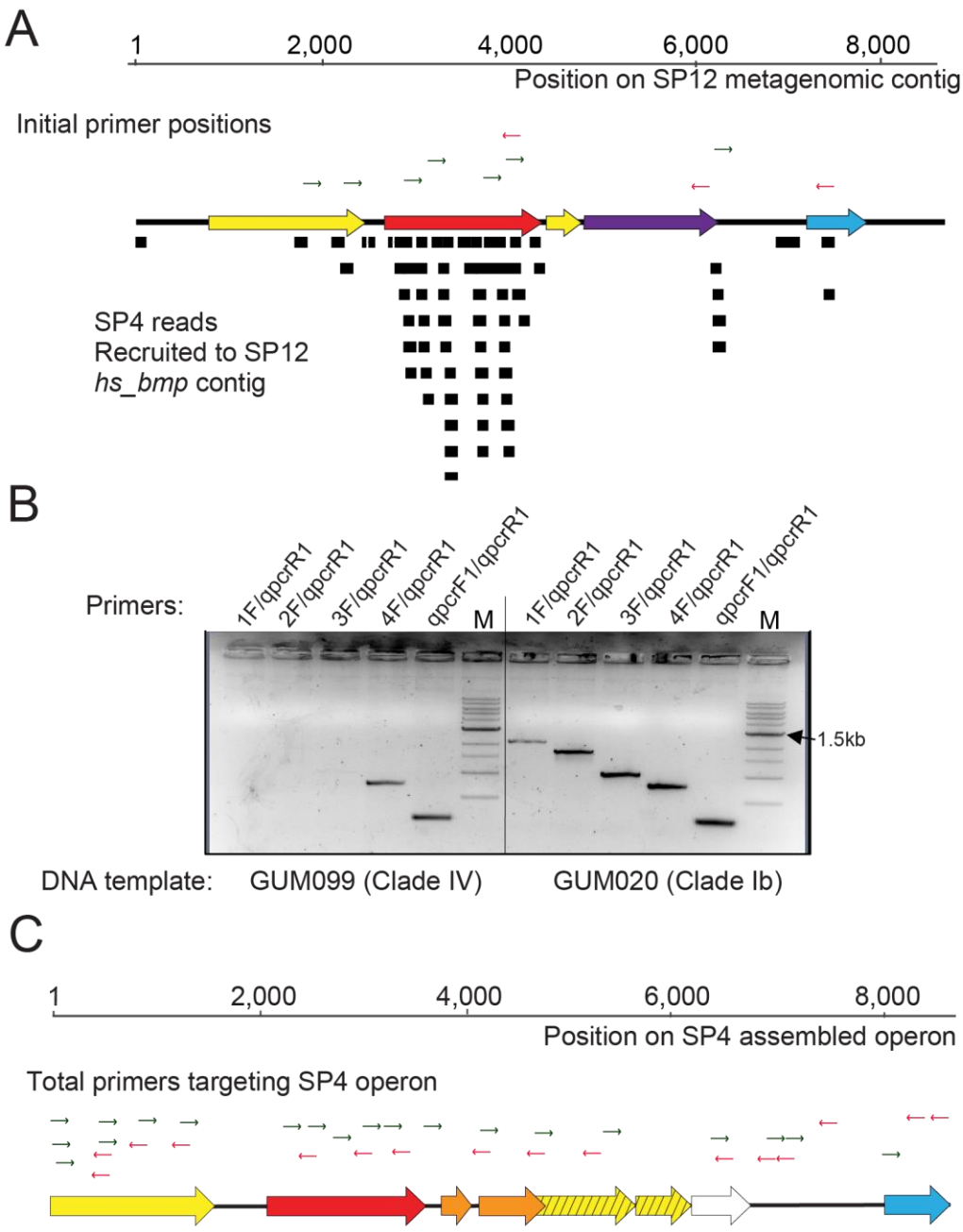
Supplementary Figure 7: Vector map of plasmid pCV0094_SP12_*hs_bmp7*-to-*hs_bmp12*. Details on the construction of this plasmid backbone are provided in the legend of Supplementary Figure 5.



Supplementary Figure 8: Metagenomic scaffold clustering for Clade IV and Clade III sponges. Plots of nucleotide composition against depth of coverage for (**A**) Clade IV sponge sample SP4 and (**B**) Clade III sponge sample GUM007. All scaffolds greater than 5 kb in length are displayed in grey. Scaffolds containing *hs_bmp* operon genes are designated by a gold circle. Points in other colors represent scaffolds classified by amino acid sequence similarity of predicted proteins to known sequences from the GenBank nr database. Note that scaffolds containing the *hs_bmp* operon were observed for the Clade IV sample but not for the Clade III sample despite the higher overall assembly coverage depth for Clade III.



Supplementary Figure 9: Quantification of PBDE levels in Clade Ia and Clade IV sponges. (**A**) Six different amounts of purified **2** from a 1 mg/mL standard solution in MeOH were analyzed by HPLC. The areas under the peaks for chromatograms recorded at 280 nm wavelength were determined by integration. The mean values with error bars are represented from three independent injections for each amount of **2** injected (3 µg, 5 µg, 10 µg, 15 µg, 20 µg, and 25 µg). Data represent mean values ± s.d. (**B**) Identical amounts of MeOH extracts of Clade Ia sponge specimen SP8 (in red) and Clade IV sample SP4 (in black) were analyzed by HPLC. Chromatograms were recorded at 280 nm wavelength and the areas under the peaks were used to dereplicate the absolute amounts of PBDEs present in each sponge specimen. Though the retention times of major PBDEs in SP8 and SP4 sponge specimens are similar, they are distinguishable by their MS1 and MS2 spectra as **5** and **2**, respectively, and have been NMR characterized previously[5].



Supplementary Figure 10: Design of primers to amplify the *hs_bmp* pathway from multiple clades. (**A**) Recruitment of SP4 (Clade IV) reads to the SP12 (Clade 1a) metagenomic contig identified potential regions of consensus, inside and outside of the *hs_bmp* ORFs (colored as in Figure 4C). Forward (green arrows) and reverse (red arrows) primers were designed to target these regions in additional clades. (**B**) Illustrative agarose gel electrophoresis analyses of PCR amplicons performed using metagenomic DNA isolated from clades IV and Ib sponge specimens with initial primers. Amplicons were Sanger sequenced to extend the known sequence in these clades, as well as to design additional primers (Supplementary Table 8). (**C**) Example of *hs_bmp* operon from SP4 with complete array of primers used for PCR amplification and sequencing.

Supplementary Tables

Supplementary Table 1. Sponge specimen collection metadata

	Host clade	Sample ID	Location	Collection year
1	Ia	SP1	Piti Bomb Holes	2014
2		SP5	Piti Bomb Holes	2014
3		SP8	Piti bomb holes	2014
4		SP12	Anae Island	2014
5		GUM038	Anae Island	2015
6		GUM040	Anae Island	2015
7		GUM058	Piti Bomb Holes	2015
8		GUM096	Piti Bomb Holes	2015
9	Ib	GUM020	Piti Bomb Holes	2015
10		GUM102	Piti Bomb Holes	2015
11	III	GUM007	Pago Bay	2015
12		GUM069	Pago Bay	2015
13	IV	GUM098	Piti Bomb Holes	2015
14		GUM099	Piti Bomb Holes	2015
15		GUM100	Piti Bomb Holes	2015
16		GUM034	Anae Island	2015
17		SP4	Piti bomb holes	2014
18		SP11	Anae island	2014

Supplementary Table 2: Assembly statistics for quality-filtered 16S rRNA Illumina MiSeq reads.

Host clade	Sample name	Number of reads
Ia	GUM038	146,821
	GUM096	143,175
Ib	GUM020	126,741
	GUM102	209,710
III	GUM007	84,657
	GUM069	132,620
IV	SP4	251,322
	SP11	74,995

Supplementary Table 3: Metagenomic assembly statistics

Host ITS Clade	Ia	IV	III
Tissue type	whole tissue	whole tissue	trichome enriched
Number input reads	45,810,804	35,453,392	8,000,000
Average read length (nt)	90	90	140
Max scaffold size	75,936	361,704	90,841
Scaffold N50*	2,321	2,877	5,876
Scaffolds > 5000 nt	3,680	4,791	480
Max cyanobacterial scaffold size	59,973	70,166	86,957
Number of 16S rRNA matches	9	31	1

*N50 is defined as the number of scaffolds required to include 50% of the total assembled nucleotides

Supplementary Table 4. Predicted genes for the *hs_bmp* operon-containing scaffold from Clade Ia, sponge specimen SP12

pathway gene annotation	start	end	ORF length (bp)	strand	protein length (aa)	Genbank match id	Match species	% aa ident	Genbank tophit lineage
hs_bmp5 halogenase	944	2533	1589	+	529	WP_008184789.1	*Moorea producens*	72	Bacteria;Cyanobacteria;Oscillatoriophycideae;Oscillatoriales;Moorea
hs_bmp7 CYP450	2818	4386	1568	+	522	WP_017308210.1	*Fischerella* sp. PCC 9339	43	Bacteria;Cyanobacteria;Stigonematales;Fischerella
hs_bmp11 fragment of *hs_bmp5* (FAD binding domain)	4516	4830	314	+	104	AFY57596.1	*Rivularia* sp. PCC 7116	81	Bacteria;Cyanobacteria;Nostocales;Rivulariaceae;Rivularia
hs_bmp12 CYPP450	4897	6258	1361	+	453	WP_006508047.1	*Xenococcus* sp. PCC 7305	64	Bacteria;Cyanobacteria;Pleurocapsales;Xenococcus
hs_bmp6 chorismate lyase	7248	7826	578	+	192	WP_054465842.1	*Planktothricoides* sp. SR001	72	Bacteria;Cyanobacteria;Oscillatoriophycideae;Oscillatoriales;Planktothricoides
transposase DDE domain protein fragment	8022	7798	75	-	48	ELS46939.1	*Microcystis aeruginosa* DIANCHI905	73	Bacteria;Cyanobacteria;Oscillatoriophycideae;Chroococcales;Microcystis
regulatory protein, tetR family	8289	8023	266	-	88	WP_015126741.1	*Calothrix* sp. PCC 7507	63	Bacteria;Cyanobacteria;Nostocales;Rivulariaceae;Calothrix
regulatory protein, tetR family	8640	8305	335	-	111	WP_016950100.1	*Anabaena* sp. PCC 7108	58	Bacteria;Cyanobacteria;Nostocales;Nostocaceae;Anabaena
DevB ABC transporter fusion protein	8893	10188	1295	+	431	WP_026735280.1	*Fischerella* sp. PCC 9605	56	Bacteria;Cyanobacteria;Stigonematales;Fischerella
DevC ABC transporter permease	10185	11363	1178	+	392	WP_026100963.1	*Synechococcus* sp. PCC 7336	63	Bacteria;Cyanobacteria;Oscillatoriophycideae;Chroococcales;Synechococcus
DevA ABC transporter ATP-binding	11482	12183	701	+	233	WP_012409968.1	*Nostoc punctiforme*	75	Bacteria;Cyanobacteria;Nostocales;Nostocaceae;Nostoc
type I restriction endonuclease	12224	12841	617	-	205	WP_015130186.1	*Calothrix* sp. PCC 7507	68	Bacteria;Cyanobacteria;Nostocales;Rivulariaceae;Calothrix
recombination protein RecR	13220	13567	347	+	115	AFY30924.1	*Calothrix* sp. PCC 7507	87	Bacteria;Cyanobacteria;Nostocales;Rivulariaceae;Calothrix
hypothetical protein	13666	14238	572	+	191	WP_019492550.1	*Calothrix* sp. PCC 7103	58	Bacteria;Cyanobacteria;Nostocales;Rivulariaceae;Calothrix
two-component response regulator	14526	14834	308	-	102	WP_009785384.1	*Lyngbya* sp. PCC 8106	61	Bacteria;Cyanobacteria;Oscillatoriophycideae;Oscillatoriales;Lyngbya
DNA (cytosine-5)-methyltransferase	14937	15320	383	+	127	AGY56851.1	*Gloeobacter kilaueensis* JS1	41	Bacteria;Cyanobacteria;Gloeobacteria;Gloeobacterales;Gloeobacter

Supplementary Table 5: Depth of coverage, homology to sequences available in Genbank database, and corresponding taxonomic lineage of 16S rRNA gene sequences recovered from the SP12 metagenome. Fields for the cyanobacterial 16S rRNA genes are highlighted.

scaffold_id	% GC	Scaffold length (nt)	Depth of coverage	Closest Genbank match	% identity	align length	Taxonomic lineage
scaffold_4453	**53**	**4108**	**113.1**	**AY615505.1**	**99.6**	**1410**	**Bacteria;Cyanobacteria;Cyanobacteria;SubsectionIII;FamilyI;Oscillatoria†;Oscillatoria spongeliae**
scaffold_4968	52	5122	18.7	JF769734.1	96.4	466	Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacterales;Hyphomonadaceae;uncultured;uncultured bacterium
scaffold_24	37	30957	14.6	AM259890.1	95.9	1460	Bacteria;Bacteroidetes;Cytophagia;Cytophagales;Flammeovirgaceae;Ekhidna;uncultured Flexibacter sp.
scaffold_33158	56	2391	12.4	AY845235.1	97.4	1381	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;uncultured;uncultured bacterium
scaffold_8961	**52**	**2642**	**10.6**	**JX206502.1**	**99.3**	**406**	**Bacteria;Cyanobacteria;Cyanobacteria;SubsectionI;FamilyI;Prochloron;uncultured bacterium**
scaffold_29473	**46**	**1410**	**6.8**	**DQ357958.1**	**97.5**	**323**	**Bacteria;Cyanobacteria;Cyanobacteria;SubsectionI;FamilyI;Prochloron;uncultured Prochloron sp.**
scaffold_42124	48	1119	6.8	JQ062806.1	92.3	545	Bacteria;Proteobacteria;Gammaproteobacteria;HOC36;uncultured bacterium

†: *Oscillatoria* is reassigned as *Hormoscilla*

Supplementary Table 6: Nucleotide/amino acid sequence identity relationships between *hs_bmp* and *bmp* genes and gene products expressed as percentage values

	bmp5*	***bmp6***	***bmp7***
hs_bmp5†	35/50		
hs_bmp6		35/54	
hs_bmp7			38/59

*****sequences from *Pseudoalteromonas luteovioloacea* 2ta16 *bmp* gene cluster[1].
†metagenomic community consensus sequences from Clade Ia SP12.

Supplementary Table 7. Heterologous expression strains used in this study

Strain name	Description
*Se*7942	*Synechococcus elongatus* PCC 7942, WT strain.
*Se*7942-C	*S. elongatus* control strain carrying at NS2 the pCV0094 kanamycin resistance marker and functional module without insert.
*Se*7942-*hs_bmp7*	*S. elongatus* carrying at NS2 the pCV0094 kanamycin resistance marker and functional module harboring *hs_bmp7*.
*Se*7942-*hs_bmp7–12*	*S. elongatus* carrying at NS2 the pCV0094 kanamycin resistance marker and functional module harboring *hs_bmp7–12*.

Supplementary Table 8: Primers used for amplification of the *hs_bmp* operon from multiple clades, with (*) denoting initial sequences designed from read recruitment of SP4 to SP12 *hs_bmp*-containing contig

Forward primer name	Sequence 5'-3'	Targets Clade Ia	Targets Clade Ib	Targets Clade IV
granulosa_bmp7_qpcrF1*	GGATCGTTGATGCAGTTTGCG	x	x	x
granulosa_bmp7_qpcrF2*	CCACTCTTCTCTGACTACTAT	x	x	x
hs_bmp_1F*	AATGCCTAACAAATGTCGATG	x	x	x
hs_bmp_2F*	TATACGGTGCTATCCAGTC	x	x	x
hs_bmp_3F*	CAGTATGTTGCCCTTCTTC	x	x	x
hs_bmp_4F*	ACCATAGTATTGTAGGTAAAGC	x	x	x
hs_bmp_5F	CTTATCAATACGGGTCTCGGAC	x	x	x
hs_bmp_10F	AGTACCATAATGCTGACAGATTTG			x
hs_bmp_11F	TTCTCTACTGTCTGGAAGCAC			x
hs_bmp_12F	TCTGGATGTATGGGAAGCMG	x	x	x
hs_bmp_14F	TATGCAGTTTAGATGAGCGACA	x	x	x
hs_bmp_18F	TAAAGTTSTAGTAAAACCTCGMTT	x	x	x
hs_bmp_19F	GAAGGTTTCGCTGCATGTC			x
hs_bmp_20F	ATGTTAGATTGCATTGTAATTGGAGC	x		x
hs_bmp_21F	TCATTCTCAGCAGTACCATAATG			x
hs_bmp_22F	TTACCGTAATGCTGACAGATTTG	x	x	
hs_bmp_22F_low	TTACCGTAATGCTGACAG	x	x	
hs_bmp_23F	TTCTACTGATGACGCTCAACA	x	x	
hs_bmp_26F	CATGTTRGATTGYATTGTAATTGGAG	x		x
hs_bmp_31F	ATGCTGGATTGTATTGTGATTGGAG	x	x	x
hs_bmp_32F	AGGAGGTTAATTTCATGCTGGATTG	x	x	x
hs_bmp_33F_low	CTTTGAATCGGCTAGTGT			x
hs_bmp_6F	TTATATTTATGCGGAGTCATTGCT	x	x	x
Reverse primer name				
granulosa_bmp7_qpcrR1*	GGGCATAGTCAGAGAAGA	x	x	x
granulosa_bmp7_qpcrR2*	TCAAGAAGGTTCTTTCTAAGT	x	x	x
hs_bmp_6R*	CGAGCAATGACTCCGCATA	x	x	x
hs_bmp_10R	TAGGGCTATGTCTGAGCCAG			x
hs_bmp_11R	AGCGCATCTGAATCAGCTC			x
hs_bmp_12R	CGGCTTCCCATACATCCAGA	x	x	x
hs_bmp_13R	ACCTCTAAGAATTCACACAATGGT		x	x
hs_bmp_16R	CAATGACTAATCCGCACCTG			x
hs_bmp_19R	GACATGCAGCGAAACCTTC			x
hs_bmp_1R	CATCGACATTTGTTAGGCATTC	x	x	x
hs_bmp_20R	GCTCCAATTACAATGCAATCTAACAT	x		x
hs_bmp_21R	AGCATTATGGTACTGCTGAGA			x
hs_bmp_22R	GCAAATCTGTCAGCATTAYGGTA	x	x	x
hs_bmp_24R	TCTACTGTATGCTGGATTAATAGCA	x	x	
hs_bmp_2R	CTGGATAGCACCGTATACC	x	x	x
hs_bmp_5R	GTCCGAGACCCGTATTGATAAG	x	x	x

Supplementary Table 9: Nucleotide-level comparison of amplified *hs_bmp* gene clusters within host clades

Samples	Clade	Alignment length (nt)	Nucleotide Percent Identity
SP12 vs GUM096	Ia	8,025	99.3
GUM020 vs GUM102	Ib	5,633	100
SP4 vs GUM034	IV	8,722	99.9

Supplementary References

1 Calcul, L. *et al.* NMR strategy for unraveling structures of bioactive sponge-derived oxy-polyhalogenated diphenyl ethers. *J. Nat. Prod.* **72**, 443-449 (2009).

2 Unson, M. D., Holland, N. D. & Faulkner, D. J. A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue. *Mar. Biol.* **119**, 1-11 (1994).

3 Utkina, N. K. *et al.* Spongiadioxins A and B, two new polybrominated dibenzo-*p*-dioxins from an Australian marine sponge *Dysidea dendyi*. *J. Nat. Prod.* **64**, 151-153 (2001).

4 Utkina, N. K., Denisenko, V. A., Virovaya, M. V., Scholokova, O. V. & Prokof'eva, N. G. Two new minor polybrominated dibenzo-*p*-dioxins from the marine sponge *Dysidea dendyi*. *J. Nat. Prod.* **65**, 1213-1215 (2002).

5 Agarwal, V. *et al.* Complexity of naturally produced polybrominated diphenyl ethers revealed via mass spectrometry. *Environ. Sci. Technol.* **49**, 1339-1346 (2015).